

一般化パレート非毎年分布の母数決定に関する一考察

A study on Parameter Estimation of Generalized Pareto Distribution

北海学園大学 工学部 ○学生員 青池 信吉
 北海学園大学 工学部 学生員 高田 和浩
 北海学園大学 工学部 正員 許士 達広

1. はじめに

近年河川整備計画において、降雨の毎年確率の適合度が低い場合の検討事項として非毎年確率が位置付けられ、非毎年確率の使用頻度が高まっている。非毎年確率を表す代表的なものに一般化パレート分布をポアソン分布で変換して年確率とする手法があり、これは分布の式が一般化極値分布と同じになる。同様に指数分布が用いられるがこれは一般化パレート分布の形状母数 $k=0$ としたものであり、非毎年の確率分布として一般化パレート分布は最も重要と言える。今回は一般化パレート分布の形状を決めるうえで重要な形状母数 k の算定方法について考察したので報告する。

2. 一般化パレート分布の非毎年データ表現

閾値を越える非毎年データの洪水ピーク値の非超過確率の分布として、一般化パレート分布が用いられており、以下のように表される。

累積分布関数

$$G(x) = 1 - \left[1 - \frac{k}{a}(x - c) \right]^{\frac{1}{k}} \quad \dots 1)$$

x : 水文データ $c = x_0$ (閾値)

$$\lambda_1 = \beta_0 = \frac{1}{M} \sum_{i=1}^M [x(i) - x_0] \quad \dots 2)$$

$$\beta_1 = \frac{1}{M(M-1)} \sum_{i=1}^M (i-1)[x_i - x_0] \quad \dots 3)$$

または

$$\beta_1 = \frac{1}{M} \sum_{i=1}^M \frac{i-0.35}{M} (x_{(i)} - x_0) \quad \dots 4)$$

$$\lambda_2 = 2\beta_1 - \beta_0 \quad \dots 5)$$

$$k = \frac{\lambda_1}{\lambda_2} - 2 \quad \dots 6)$$

$$a = \lambda_1(1+k) \quad \dots 7)$$

ただし $x_i - x_0 > 0$ である。

2) ~ 4) 式は確率重みつき積率 (PWM) であり、2) 式は 0 次積率で平均に等しく、3) 式は 1 次積率に 2 項分布を利用したもの、4) 式は同じくプロットングポジションを利用した例である。いずれも i が 1 の時には $x_{(i)}$ が最小、 i が M のときには最大である。

6) ~ 8) 式はそれを用いて定数を L 積率で定める場合の式である。1) 式及び 6) 式の k が今回算定法を問題にする定数(形状母数)である。 k は算出法によって値が異なり、それが確率値に影響する。 k は一般化パレート分布から一般化極値分布に変換しても変化しないので、通常一般化パレート分布から 2 ~ 6) 式で求められる。

3. 分布のプロットングポジションによる表現

非毎年の場合の k 値推定法は重みつき積率法が主に紹介されているが、他の方法は知られていない。ここでは、積率法としてひずみ係数合わせを考えてみる。非毎年の水文データ x を y 軸、分布の標準変数 ε を x 軸に取りプロットしてみる。通常のプロットングポジションによる非超過確率 p は以下の式であり

$$p_i = \frac{i - \alpha}{M + 1 - 2\alpha} \quad \dots 11)$$

一般化パレート分布では下から i 番目の確率値 x_{pi} は

$$\begin{aligned} x_{pi} &= c + \left(\frac{a}{k} \right) \left[1 - (1 - p_i)^k \right] \\ &= c + a \left[\frac{1}{k} \left\{ 1 - \left(1 - \frac{i - \alpha}{M + 1 - 2\alpha} \right)^k \right\} \right] \quad \dots 12) \end{aligned}$$

α : プロットングポジションの定数

p : 非超過確率 M : 閾値以上のデータ個数

i : プロットングポジション

a, k, c : 分布の母数 (定数) $c = x_0$ (閾値)

なおポアソン変換したプロットングポジションによる一般化極値分布の確率値は以下の式のようになり、これからも歪み係数合わせで同じ k 値が求められる。

$$\begin{aligned} x_{pi} &= c + a \left[\frac{1}{k} \left\{ 1 - \{-\ln(p)\}^k \right\} \right] \\ &= c + a \left[\frac{1}{k} \left\{ 1 - \left\{ \frac{M}{N} \left(1 - \frac{i - \alpha}{M + 1 - 2\alpha} \right) \right\}^k \right\} \right] \quad \dots 13) \end{aligned}$$

N : データ年数 λ : ポアソン分布の閾値以上の値の年平均生起回数

[] 内が標準変数 ε_i であり、

$$x_p = a * \varepsilon_i + c \quad \dots 14)$$

の直線関係となる。

4. 母数 k の決定方法

1) PWM による方法

現在一般的に用いられている方法であり、一般化パレート分布から前出の 2)～7)式を用いて重みつき積率 PWM と L 積率を組み合わせる算出される。β₁の算出方法が何通りもあり、2項分布を用いた 3)式、PWM のプロットングポジションとして多く用いられる 4)式

$$\text{がある。このほか 4) 式の } \frac{i-0.35}{M} \cdots 15)$$

の代わりに一般的なプロットングポジション 11)式を用いることが考えられる。また、国際的に知られるハンドブックにおいてはプロットングポジションを

$$1 - \frac{i-0.35}{M} \cdots 16)$$

として i = 1 ~ M に対しデータは大きい順に対応させる、すなわち x₍₁₎ ≥ ... ≥ x_(M) として β₁を算出する方法が示されている。

2) ひずみ係数あわせによる k 値の算定

データ値と分布の標準変量が直線関係にあることから、17) 式の関係を用いて標本ひずみ係数を求めれば、以下

の関係式が求まる。両辺の値にデータ数の平方根 √M をかければ、標本ひずみ係数 C_sになる。これはいわゆる積率解であり、ここでは「歪み係数合わせ」と呼ぶことにするが、確率値と標準変量が直線関係で表される分布全てに適用できる。

$$\frac{\sum_{i=1}^M (x_i - \bar{x})^3}{\left\{ \sum_{i=1}^M (x_i - \bar{x})^2 \right\}^{3/2}} = \frac{\sum_{i=1}^M (\varepsilon_i - \bar{\varepsilon})^2}{\left\{ \sum_{i=1}^M (\varepsilon_i - \bar{\varepsilon})^2 \right\}^{3/2}} \cdots 17)$$

x_i : 下から i 番目の水文データ値、ε_i : 一般化極値

分布の標準変量

M : 閾値以上の対象データ数

N年間から M 個のデータを取り出した場合、k を仮定し

て 12) 式の ε_i を 17) 式右辺に代入し、その時の実際の

データによる左辺と一致するように k をトライアルで求める。なお 13) 式を 17)式に代入して k をトライアルで求めても、求まる k 値は同じである。

3) 理論解による方法

一般化パレート分布の理論解は

$$\gamma_x = \frac{2(1-k)(1+2k)^{1/2}}{(1+3k)} \cdots 18)$$

でしめされる。左辺は実際のデータによる不偏ひずみ係数であり、左辺の値に一致するように k をトライアルで求める。

5. 計算法の比較

k 値算定にどの手法が適当か、母数 a、c、k を仮定して、12) 式における標準変量の中の i を変化させて一般化パレート分布のデータを発生させ、前章のそれぞれの方法で算定した k 値と、発生のために仮定した k 値が一致するかどうかを確認した。

結果は表 1 に示すが、表の算出 k 値のうち、2 項は 3) 式、プロット 1 は 11) 式、プロット 2 は 16) 式、逆プロットは 17) 式、歪み係数は 19) 式のものを意味する。表の左側に示す仮定した k 値に近い k 値を算出する方法が精度が高いということになる。なお a、c の値を代えて計算したが、算出される k 値は a、c の値を変化させても変わらない。また 18) 式の歪み係数あわせでは仮定した k 値と全く同じ k 値が算出される。

表 1 k 値の算出

データ数 M	発生 k 仮定	算出 k 値				
		2項	プロット1	プロット2	逆プロット	歪み係数
30	-0.12	-0.124	-0.046	-0.059	-0.021	0.013
60		-0.118	-0.080	-0.086	-0.068	-0.017
120		-0.117	-0.098	-0.101	-0.092	-0.042
300		-0.117	-0.110	-0.111	-0.107	-0.068
3000		-0.119	-0.119	-0.119	-0.118	-0.102
30	-0.25	-0.232	-0.159	-0.170	-0.137	-0.046
60		-0.233	-0.197	-0.203	-0.187	-0.083
120		-0.237	-0.219	-0.222	-0.214	-0.115
300		-0.241	-0.234	-0.235	-0.232	-0.148
3000		-0.248	-0.247	-0.247	-0.247	-0.198
30	-0	-0.021	0.061	0.047	0.090	0.080
60		-0.009	0.031	0.025	0.045	0.059
120		-0.004	0.016	0.013	0.023	0.041
300		-0.002	0.006	0.005	0.009	0.024
3000		-0.001	0.000	0.000	0.000	0.005

これで見ると全ての場合で発生データ数を大きくするほど k 値はデータ発生に用いたものに近づく。重みつき積率法で適合度が高いのは 2 項分布を用いたものであり、プロットングポジションの中では、(i-0.35)/M がやや成績が良く、実用的に重みつき積率で多く使われていることを実証した形になっている。なお 18) 式の歪み係数合わせの方法は仮定した値を再現するが、今回のような分布どおりに発生させたデータではない、実際の観測データで検証する必要がある。

あとがき

重み付き積率では 2 項分布を用いたものが良好な結果を与えるということが分かったが、歪み係数あわせとどちらが良いかの判断はこれからである。非毎年の年確率については、指数分布のポアソン変換と一般化パレート分布を変換した一般化極値分布の値が大きく異なるなど課題が多い。特に k 値の算定方法が重要と考えられ、今後も検討が必要である。

参考文献

星清：洪水ピークの確率評価法について、開発土木研究所月報、№.539、p34 - 47、1998 年 4 月
Jery R.Stedinger: Frequency Analysis of Extreme Events , Handbook of Hydrology Chpter18 1993